



Ciencia de Datos

FACULTAD DE MATEMÁTICA Y COMPUTACIÓN

# INTRODUCCIÓN A LA CIENCIA DE DATOS PROYECTO FINAL

*Procesos migratorios en Cienfuegos*

AUTOR:  
LUIS ERNESTO SERRAS RIMADA



# 1 Introducción

La búsqueda de mejores oportunidades de empleo y educación es un impulso fundamental que motiva a las personas a migrar, tanto a nivel interno como externo. Este fenómeno refleja no solo la aspiración individual de alcanzar una vida más digna y satisfactoria, sino también un contexto socioeconómico que, en muchos casos, limita el desarrollo personal y profesional en sus lugares de origen. La decisión de emprender este camino, aunque a menudo dolorosa y compleja, simboliza la lucha por la superación y el deseo de construir un futuro mejor. Al mismo tiempo, pone de manifiesto las disparidades entre regiones, donde lugares como [La Habana](#) ofrecen un abanico más amplio de oportunidades comparado con provincias como [Cienfuegos](#). En última instancia, este deseo de progreso se convierte en una fuerza dinámica que puede transformar realidades, tanto para los individuos como para las comunidades a las que pertenecen.

## 2 Desarrollo

### 2.1 Recopilación y procesamiento de los datos

La fuente principal de los datos manejados fueron las series estadísticas y anuarios de la [Oficina Nacional de Estadísticas e Información \(ONEI\)](#). Las series estadísticas empleadas en formato .xls (llevados a formato csv) fueron reacomodados a mano mayormente para luego limpiarlos haciendo uso del modulo `csv` de [Python](#).

El análisis en su totalidad fue realizado manejando datos procesados y cargados unicamente desde la manipulación directa de los `csv` por medio del módulo de [pandas](#).

```
1 # Function to skip initial space on csv's files
2 def skipinitialspace_csv(path:str):
3     from csv import reader, writer
4     data = []
5     with open(path, "r") as f:
6         reader = reader(f, skipinitialspace=True)
7         for i in reader:
8             data.append(i)
9     with open(path, 'w') as f:
10         writer = writer(f)
11         writer.writerows(data)
```

Listing 1: Funcion para la limpieza de los csv (eliminar espacios adicionales)

```
1 def migratory_movements(df: pd.core.frame.DataFrame, type:str="prov") -> list[
2     pd.core.frame.DataFrame]:
3     #Funcion para el procesamiento de los datos
4     if type == "prov":
5         year2012, year2013, year2014, year2015, year2016, year2017, year2018,
6         year2019, year2020, year2021, year2022 = 0,0,0,0,0,0,0,0,0,0,0,0
7         dfs = [year2012, year2013, year2014, year2015, year2016, year2017, year2018
8         , year2019, year2020, year2021, year2022]
9         i = 1
10        for index in range(len(dfs)):
11            j = i + 49
12            dfs[index] = df.iloc[i:j,:].
13            dfs[index].set_index("PROCEDENCIA/DESTINO", inplace=True)
14            dfs[index].index.name='Anno'
15            i = j + 3
16        elif type == "mun":
17            year2019, year2020, year2021, year2022 = 0,0,0,0
18            dfs = [year2019, year2020, year2021, year2022]
19            i = 1
20            for index in range(len(dfs)):
21                j = i + 8
22                dfs[index] = df.iloc[i:j,:].
23                dfs[index].set_index("ANNOS", inplace=True)
24                dfs[index].index.name='Anno'
25                i = j + 1
26        elif type == "etc":
27            prov1, prov2, prov3, prov4, prov5, prov6, prov7, prov8, prov9, prov10,
28            prov11, prov12, prov13, prov14, prov15 = 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
```

```

25     dfs = [prov1, prov2, prov3, prov4, prov5, prov6, prov7, prov8, prov9,
26     prov10, prov11, prov12, prov13, prov14, prov15]
27     i = 1
28     for index in range(len(dfs)):
29         j = i + 12
30         dfs[index] = df.iloc[i:j,:].
31         dfs[index].set_index("PROVINCIA/ANNOS", inplace=True)
32         dfs[index].index.name='Anno'
33         i = j + 1
34     return dfs

```

Listing 2: Funcion utilizada para parte del procesamiento de algunos datos

Para manejar los datos se hizo uso de la historia de [Perla](#), una joven cienfueguera que (como parte de su desarrollo como estudiante de 12mo grado de Preuniversitario, situación familiar y residencial) se enfrenta a los dilemas que se desean exponer, y de esa forma según se avanza en la historia se van presentando los datos visualizandolos de forma natural e interactiva siguiendo la historia.

Y finalmente para el mapa de densidad se trabajó con la fuente del repositorio de [geojsons](#) de Cuba de [Yudivián Almeida](#). A partir de ellos, además de establecer las divisiones del territorio nacional y por provincias, se fueron agregando los valores que se quisieron representar en el mapa y en el tooltip por medio del siguiente script:

```

1     import json
2     import pandas as pd
3     '''
4     Script para agregar los valores que se mostrarían en el tooltip del mapa al
5     geojson
6     '''
7     with open("app/data/geojsons/cuba.geojson") as json_file:
8         data = json.load(json_file)
9         path = "app/data/csv/Movimiento migratorio interno por sexos y provincias.csv"
10
11     def skipinitialspace_csv(path:str):
12         from csv import reader, writer
13         data = []
14         with open(path, "r") as f:
15             reader = reader(f, skipinitialspace=True)
16             for i in reader:
17                 data.append(i)
18         with open(path, 'w') as f:
19             writer = writer(f)
20             writer.writerows(data)
21         skipinitialspace_csv(path)
22
23     df = pd.read_csv(path)
24     def migratory_movements(df: pd.core.frame.DataFrame) -> list[pd.core.frame.
25     DataFrame]:
26         year2012, year2013, year2014, year2015, year2016, year2017, year2018,
27         year2019, year2020, year2021, year2022 = 0,0,0,0,0,0,0,0,0,0,0,0
28         dfs = [year2012, year2013, year2014, year2015, year2016, year2017, year2018
29         , year2019, year2020, year2021, year2022]
30         i = 1
31         for index in range(len(dfs)):
32             j = i + 49
33             dfs[index] = df.iloc[i:j:3,:].
34             dfs[index].set_index("PROCEDENCIA/DESTINO", inplace=True)
35             dfs[index].index.name=None
36             i = j + 3
37         return dfs
38
39     df = migratory_movements(df)
40     years = [x for x in range(2012,2023)] #Ordered list of years
41     provincias = list(df[0].columns)[1:]
42
43     #Provincias en el orden que aparece en la estructura del GeoJson
44     lista_prov = ["La Habana","Matanzas","Cienfuegos","Sancti Spiritus","Las Tunas"
45     ,"Holguin","Granma","Santiago de Cuba","Isla de la Juventud"]

```

```

41         , "Camaguey", "Ciego de Avila", "Villa Clara", "Guantanamo", "Pinar
del Rio", "Artemisa", "Mayabeque"]
42
43     index = 0
44     n = list(df[0].index)
45     for year in years:
46         for index, prov in enumerate(lista_prov):
47             values = list(df[years.index(year)].iloc[:, n.index(prov)])[1:]
48             for k, j in list(zip(provincias, values)):
49                 data["features"][index]["properties"][f"{k}{year}"] = j
50
51     try:
52         with open("app/data/geojsons/cuba.geojson", 'w') as archivo_json:
53             json.dump(data, archivo_json, indent=4)
54             print("GeoJson Actualizado de forma exitosa")
55     except Exception as e:
56         print(f"Error al procesar el GeoJson: {e}")

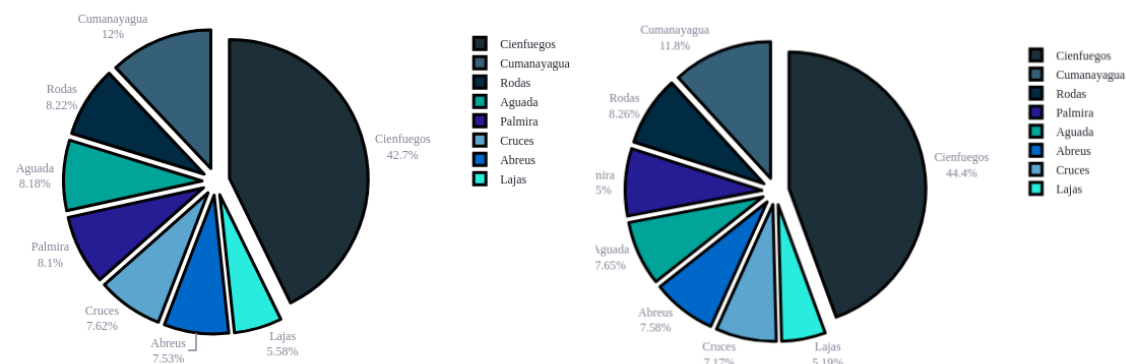
```

Listing 3: Script para agregar valores al geojson

## 2.2 Visualizaciones

Como parte principal de este análisis se procede a visualizar los datos recopilados para valorar su comportamiento.

### 2.2.1 Población

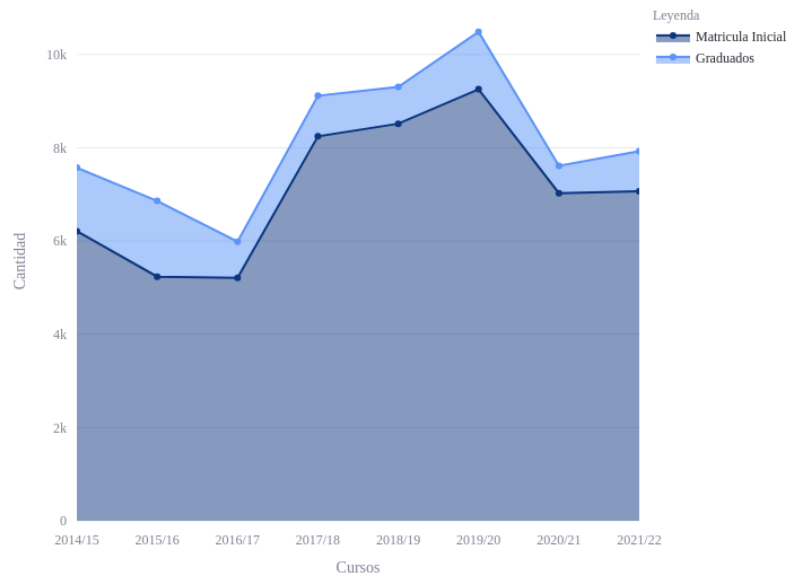


Se desea partir de analizar la distribución territorial del municipio que se quiere estudiar, por lo que con esta representación ([dividida por edad no laboral a la izquierda](#) y [edad laboral a la derecha](#)) se descubre el predominio de densidad del municipio cabecera con respecto al resto reflejando un claro desblanace de gestión de recursos municipales.

### 2.2.2 Educación superior

Siguiendo el devenir de la narrativa expuesta se introduce una observación referente al número de matriculas iniciales y graduados en comparación para cada año en [Cienfuegos](#) de la educación superior (la cual cabe resaltar que sólo cuenta con 3 universidades y todas en el municipio cabecera)

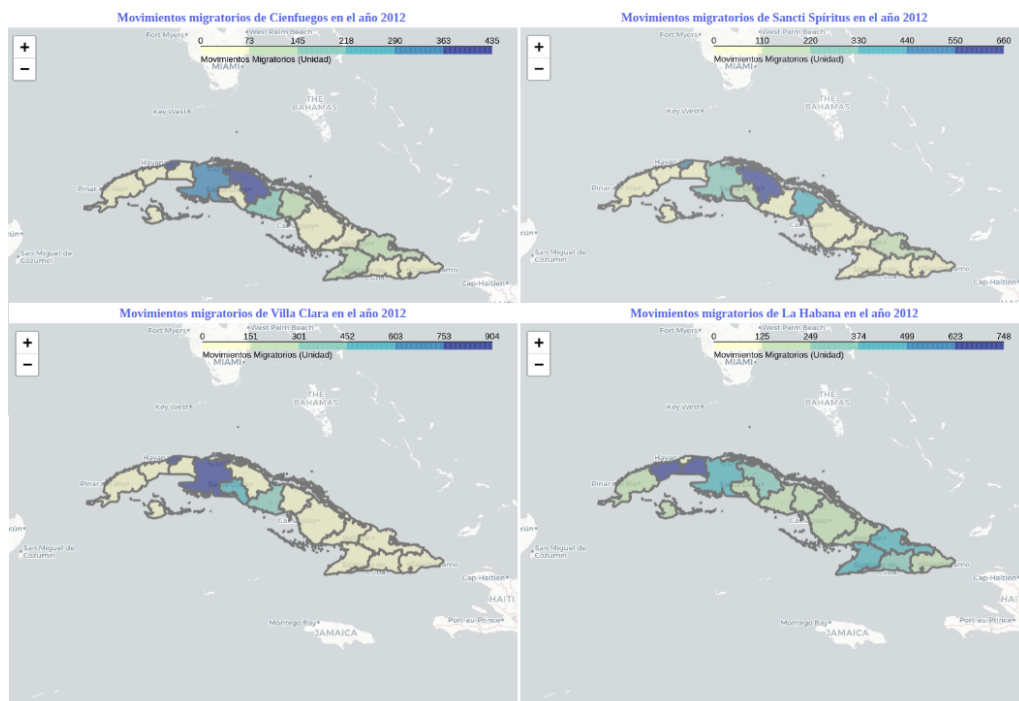
Alcanzando a reflejarse el poco volumen de matrículas y la enorme diferencia referente al número de graduados para cada año en una provincia cuyas únicas instituciones de educación superior radican en la cabecera, por lo que claramente las características del entorno para el escenario de quedarse en su provincia natal van esfumando toda idea o interés por seguir en ese sitio. En medio de esta problemática nace la problemática de la toma de una decisión crucial para la chica, estudiar en la capital lo que quiere o conformarse con lo que su municipio le ofrece, cada alternativa con sus características particulares descritas.



Luego, acto seguido a introducir al personaje del padre de la joven con su acción de migrar en el pasado, se procede a analizar los procesos migratorios del territorio cubano y del municipio en profundidad.

### 2.2.3 Movimientos migratorios

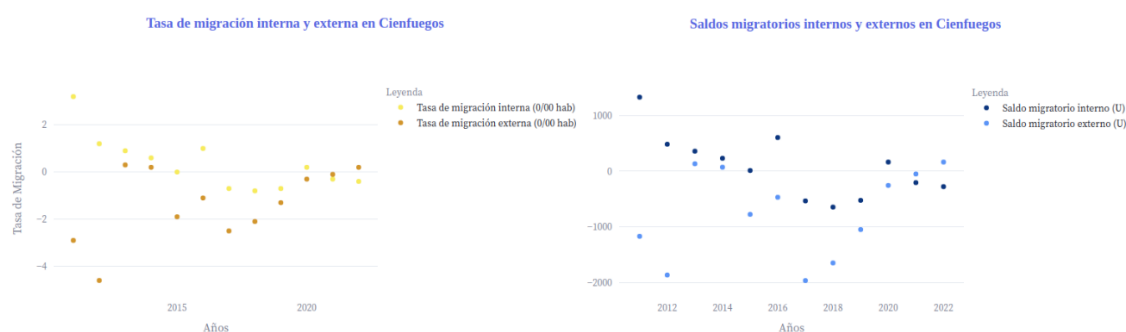
Dentro de este apartado se plantea estudiar los movimientos migratorios en la isla por provincias para así evaluar de forma nacional y realizar comparaciones aprovechando la interactividad del recurso presentado en un mapa de densidad.



Se observa que los flujos más significativos se dirigen principalmente hacia las provincias de **Villa Clara**, **Matanzas** y **La Habana**. Por otro lado hacia **Cienfuegos** se tiene, en muy pocos volúmenes, a provincias como **Sancti Spíritus**, **La Habana**, **Matanzas** y en mayores valores con **532 unidades para el año 2012** se encuentra **Villa Clara** como principal origen.

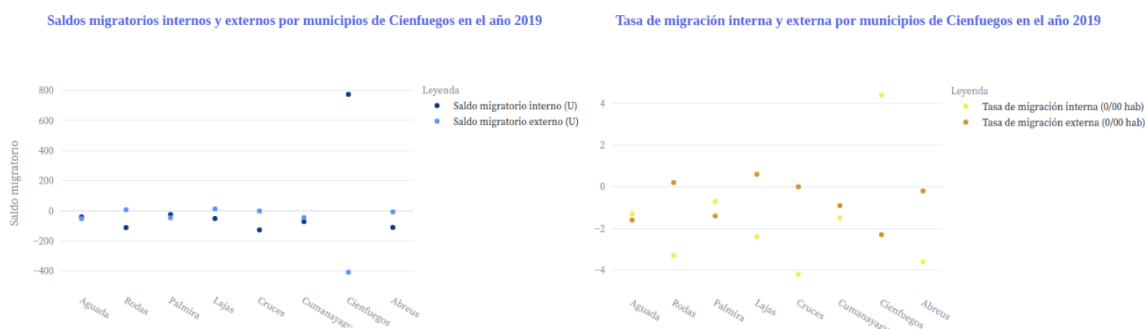
## 2.2.4 Saldo y tasa de migración

Luego, se aprecia que a diferencia de otras regiones de Cuba, **Cienfuegos** presenta una notable estabilidad en su saldo migratorio interno, lo que significa que los movimientos migratorios interprovinciales se equilibran de manera más favorable en comparación con el típico saldo negativo que caracteriza al país en su conjunto de forma general.



Sin embargo, al analizar la dinámica migratoria a nivel municipal, se evidencia una tendencia significativa hacia el municipio de cabecera, **Cienfuegos**, dominando la densidad poblacional del municipio agrupando a más del 40% de la población residente. Este dato revela que, a pesar de la estabilidad general de **Cienfuegos**, existe un posible desbalance intermunicipal que podría estar impulsado por la búsqueda de mejores oportunidades laborales, educación y calidad de vida.

Además, al considerar los diferentes tipos de saldos migratorios, se concluye que, en términos generales, el saldo externo es el que se lleva la delantera con valores negativos (excluyendo la diferencia de saldo interno del municipio **Cienfuegos**). Esto sugiere que, si bien **Cienfuegos** mantiene un equilibrio interno más sólido, la migración hacia el extranjero también juega un papel crucial en la configuración de su demografía y, por ende, en el futuro desarrollo de la región.

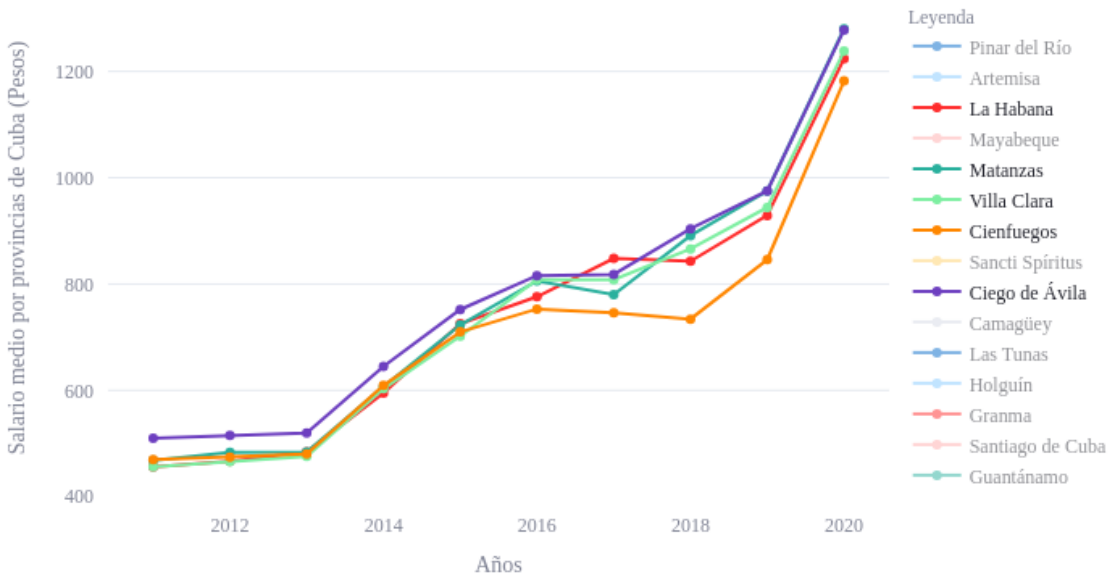


## 2.2.5 Salario medio y empleos

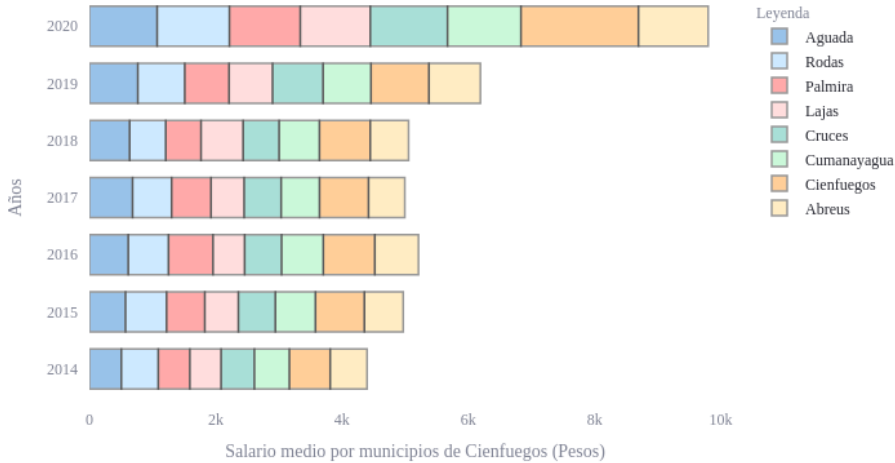
En el ámbito de las oportunidades laborales (desarrollando en base al contexto de la narración en busca de intentar explicar o hallarle un motor a de Andrés, el padre de Perla, de abandonar a su familia), se tiene a **La Habana** se posiciona como un imán para quienes buscan una mejor calidad de trabajo y un nivel de vida más elevado. En comparación con **Cienfuegos**, la capital ofrece mayores posibilidades de empleo, con una variedad más amplia de sectores y empresas que garantizan salarios medios más altos. Además, en **La Habana**, es más común acceder a créditos y a una gama de servicios que facilitan el desarrollo personal y profesional, lo que atrae a muchos migrantes de otras provincias en busca de un futuro más prometedor.

Entonces se decide evaluar (con el fin de minimizar el número de líneas, excluyendo las que no nos

aporten al análisis siguiendo el hilo de desarrollo), considerando únicamente las provincias que se conoce que imperan como mayoría como destino de las migraciones desde Cienfuegos, quedándose por debajo con respecto al resto de provincias.



De igual forma se puede decir, si consideramos las diferencias intermunicipales, que se mantienen con ligera superioridad general los valores de media salarial para la cabeza del municipio, resaltando una vez más el desbalance estructural de distribución de recursos municipales existente en dicho territorio.



*Cabe mencionar que aunque ciertamente esta representación a priori puede que no presente un caracter visualmente significativo, se considera emplear este recurso y la interactividad del mismo para aprovechar el formato de pila barras y reflejar las diferencias de cada municipio con respecto al resto (y de forma general también el desarrollo anual del salario medio del municipio entre los considerados).*

### 3 Encuesta

Con el objetivo de recopilar perspectivas sobre el tema en cuestión, se llevó a cabo una encuesta online utilizando la aplicación forms.app, la cual fue realizada por 45 personas. Aunque es importante señalar que los resultados de esta encuesta no poseen valor estadístico general debido a su carácter no representativo, no obstante la cantidad de participantes y la estructura intuitiva del cuestionario fomentaron interacciones valiosas y significativas.

La encuesta fue diseñada para facilitar la participación y animar a los encuestados a compartir sus opiniones, lo que permitió obtener una variedad de respuestas que enriquecen el análisis del tema. A pesar de no ser un conjunto de datos estadísticamente válido, la diversidad de voces y experiencias recogidas brinda un panorama interesante que puede ser considerado para profundizar en la discusión.

- El desarrollo del formulario planteado sigue el siguiente flujo: <https://github.com/LFrench03/La-Perla-del-Sur/tree/main/app/form>
- Para ver más detalles (audiencia, respuestas y opiniones) ver en la web: <https://la-perla-del-sur.streamlit.app/>

### 4 Conclusiones y recomendaciones

En conclusión, el estudio de los procesos migratorios en [Cienfuegos](#) es crucial para comprender las dinámicas de movilidad poblacional y sus repercusiones socioeconómicas. Las observaciones revelan que la migración hacia la cabecera municipal está motivada por factores como la búsqueda de empleo, acceso a servicios y mayores oportunidades educativas. Sin embargo, esta concentración demográfica también genera desafíos en términos de infraestructura y sostenibilidad. Se recomienda implementar políticas que no solo fortalezcan la infraestructura de la cabecera, sino que también promuevan el desarrollo de las áreas periféricas, garantizando la equidad en el acceso a recursos y oportunidades. Asimismo, es esencial establecer mecanismos de monitoreo para evaluar continuamente el impacto de la migración y ajustar las estrategias en función de las necesidades cambiantes de la población.